JOURNAL OF COMBINATORIAL THEORY, Series A, **79**, 173–179 (1997)
ARTICLE NO. TA962773

NOTE

Urn Sampling and a Majorization Inequality

Kiran S. Kedlaya

Department of Mathematics, Princeton University, Princeton, New Jersey 08544

Communicated by the Managing Editors

Received August 16, 1996

conjectured that the sequences $\mathbf{x}_n = \binom{A-n}{a}$ and $\mathbf{y}_n = \binom{B-n}{b}$ form such a pair for any nonnegative integers $A \geqslant a$, $B \geqslant b$, and proved this result in the cases $\min\{A, B\} \geqslant a+b$ and $-b \leqslant A-B \leqslant a$. We complete the proof of the conjecture by proving the result under the assumption $\max\{A, B\} \geqslant a+b$.

1. INTRODUCTION

The inequality studied in this paper arises from the following game introduced by Bennett [2]. Let A, a, B, b be nonnegative integers with $\min\{A, B\} \geqslant a+b$, and let N be a positive integer. Consider two urns, the first containing a black balls and $A+1-a$ white balls, the second containing b black balls and $B+1-b$ white balls. We repeatedly select one ball at random from each urn and set the selected balls aside, until one urn is exhausted.

Two players wager independently on the outcome of this experiment as follows. Alfie selects N distinct pairs of positive integers and wins one dollar if the first white ball drawn from the first urn occurs on the mth draw and the first white ball drawn from the second urn occurs on the nth draw, where (m, n) is one of his chosen pairs. Betty selects N distinct positive integers and wins one dollar if the kth draw is the first time two white balls are chosen, where k is one of her chosen numbers. (The condition $\min\{A, B\} \geqslant a+b$ ensures that a black ball cannot appear on every draw.)

The outcome of this game, as demonstrated in [2] and summarized in the following theorem, may be a bit surprising.

0097-3165/97 $25.00

THEOREM 1. *For any given N, if both players make selections so as to maximize their expected payoffs, the expected payoff for Alfie is no greater than the expected payoff for Betty.*

We now introduce some notation we will use to restate the result in an algebraic form, and ultimately to prove a stronger result. To any 4-tuple (A, a, B, b) of nonnegative integers with $A \geqslant a$ and $B \geqslant b$, we associate the infinite sequences

$$\mathbf{x}_n = \binom{A-n}{a}, \qquad \mathbf{y}_n = \binom{B-n}{b}. \tag{1}$$

The difference operator Δ acts on sequences as follows:

$$\Delta \mathbf{x}_n = \mathbf{x}_n - \mathbf{x}_{n+1}.$$

We write $\Delta^0 \mathbf{x} = \mathbf{x}$ and $\Delta^m \mathbf{x} = \Delta(\Delta^{m-1}\mathbf{x})$. By induction, we see that for all $m \geqslant 0$,

$$\Delta^m \mathbf{x}_n = \sum_{k=0}^{m} (-1)^k \binom{m}{k} \mathbf{x}_{n+k}.$$

We also write $\mathbf{xy}$ for the sequence whose nth term is $\mathbf{x}_n \, \mathbf{y}_n$.

In this notation, the above theorem states that under the assumption $a + b \leqslant \min\{A, B\}$, the double sequence $\{(\Delta^m \mathbf{x}_0)(\Delta^n \mathbf{y}_0)\}_{m,n=0}^{\infty}$ is majorized by the single sequence $\{\Delta^k(\mathbf{xy})_0\}_{k=0}^{\infty}$. (In other words, for any N, the sum of the N largest terms of the double sequence does not exceed the sum of the N largest terms of the single sequence.) In Bennett's terminology, the sequences $\mathbf{x}$ and $\mathbf{y}$ are said to form a *double-dipping pair*.

Bennett conjectured that this majorization occurs under the weaker conditions $A \geqslant a$ and $B \geqslant b$; in [3], he proved this conjecture in the case $-b \leqslant A - B \leqslant a$. We shall verify the conjecture in the case $a + b \leqslant \max\{A, B\}$. This completes the proof of the conjecture, since if $a + b > A$ and $a + b > B$, then $A - B < a + b - B \leqslant a$ and similarly $B - A < b + a - A \leqslant b$.

In fact, for $a + b \leqslant \max\{A, B\}$, we prove the stronger result that for any N, the sum of any N terms of the double sequence $\{(\Delta^m \mathbf{x}_0)(\Delta^n \mathbf{y}_0)\}_{m,n=0}^{\infty}$ does not exceed the sum of the *first* N terms of the sequence $\{\Delta^k(\mathbf{xy})_0\}_{k=0}^{\infty}$. In symbols, for any distinct ordered pairs (c_i, d_i) $(i = 0, ..., N-1)$ of nonnegative integers,

$$\sum_{i=0}^{N-1} \Delta^i(\mathbf{xy})_0 \geqslant \sum_{i=0}^{N-1} (\Delta^{c_i} \mathbf{x}_0)(\Delta^{d_i} \mathbf{y}_0). \tag{2}$$

It should be noted that this stronger inequality need not hold under the weaker conditions $A \geqslant a$ and $B \geqslant b$. For example, if $A = a = B = b = 2$, it is easily checked that the inequality fails for $N = 4$. This failure may explain why Bennett's proof for the case $-b \leqslant A-B \leqslant a$ is more intricate than the argument in this paper.

2. RESULTS

LEMMA 1. *For any sequences* $\mathbf{x}$ *and* $\mathbf{y}$,

$$\Delta^n(\mathbf{xy})_k = \sum_{j=0}^{n} \binom{n}{j} \Delta^{n-j}\mathbf{x}_{k+j}\, \Delta^j \mathbf{y}_k.$$

Proof. We induct on n. The case $n=1$ is the obvious fact that

$$\mathbf{x}_k \mathbf{y}_k - \mathbf{x}_{k+1}\mathbf{y}_{k+1} = (\mathbf{x}_k - \mathbf{x}_{k+1})\, \mathbf{y}_k + \mathbf{x}_{k+1}(\mathbf{y}_k - \mathbf{y}_{k+1}).$$

For $n>1$, we assume the result for $n-1$ and find

$$\begin{aligned}
\Delta^n(\mathbf{xy})_k &= \sum_{j=0}^{n-1} \binom{n-1}{j} \Delta(\Delta^{n-1-j}\mathbf{x}_{k+j}\, \Delta^j \mathbf{y}_k) \\
&= \sum_{j=0}^{n-1} \binom{n-1}{j} (\Delta^{n-j}\mathbf{x}_{k+j}\, \Delta^j \mathbf{y}_k + \Delta^{n-1-j}\mathbf{x}_{k+j+1}\, \Delta^{j+1}\mathbf{y}_k) \\
&= \sum_{j=0}^{n-1} \binom{n-1}{j} \Delta^{n-j}\mathbf{x}_{k+j}\, \Delta^j \mathbf{y}_k + \sum_{i=1}^{n} \binom{n-1}{i-1} \Delta^{n-i}\mathbf{x}_{k+i}\, \Delta^i \mathbf{y}_k \\
&= \sum_{j=0}^{n} \binom{n}{j} \Delta^{n-j}\mathbf{x}_{k+j}\, \Delta^j \mathbf{y}_k. \quad \blacksquare
\end{aligned}$$

In the case $\mathbf{x}_n = \binom{A-n}{a}$, $\mathbf{y}_n = \binom{B-n}{b}$, the lemma implies that

$$\Delta^n(\mathbf{xy})_0 = \sum_{j=0}^{n} \binom{n}{j} \binom{A-n}{a-n+j} \binom{B-j}{b-j}.$$

We call the 4-tuple (A, a, B, b) *acceptable* if (2) holds for the sequences defined in (1).

LEMMA 2. *For any nonnegative integers* a *and* b, $(a+b, a, b, b)$ *is acceptable.*

Proof. For $0 \leqslant n \leqslant a+b$, we have

$$\begin{aligned}
\Delta^n(\mathbf{xy})_0 &= \sum_{j=0}^{n} \binom{n}{j} \binom{a+b-n}{a-n+j} \binom{b-n}{b-n} \\
&= \sum_{j=0}^{b} \binom{n}{j} \binom{a+b-n}{b-j} = \binom{a+b}{b},
\end{aligned}$$

where the last step is a Vandermonde convolution. Therefore the first $a+b+1$ terms of the sequence $\Delta^k(\mathbf{xy})_0$ are all equal to $\binom{a+b}{b}$, and subsequent terms are zero. On the other hand, the nth largest term of the double sequence is $\binom{a+b-t}{a-t}$, where $t=\lfloor (n-1)/(b+1)\rfloor$. This is no greater than $\binom{a+b}{b}$; moreover, the sum of the entire double sequence is $(b+1)\binom{a+b+1}{a}=(a+b+1)\binom{a+b}{b}$, the same as the sum of the single sequence. The desired inequality is now clear. ∎

LEMMA 3. *For any nonnegative integers A, a, b such that $A \geqslant a+b$, (A, a, b, b) is acceptable.*

Proof. We proceed by induction on $A-a-b$, with the base case $A-a-b=0$ following from the previous lemma. Within this induction, we perform a second induction on a, with trivial base case $a=0$.

By the induction hypothesis on $A-a-b$, $(A-1, a, b, b)$ is acceptable, while by the induction hypothesis on a, $(A-1, a-1, b, b)$ is acceptable. This gives us a pair of inequalities for any N:

$$\sum_{n=0}^{N-1}\left(\sum_{j=0}^{b}\binom{n}{j}\binom{A-1-n}{a-n+j}\right)\geqslant\sum_{n=0}^{N-1}\binom{A-1-\lfloor n/(b+1)\rfloor}{a-\lfloor n/(b+1)\rfloor}$$

$$\sum_{n=0}^{N-1}\left(\sum_{j=0}^{b}\binom{n}{j}\binom{A-1-n}{a-1-n+j}\right)\geqslant\sum_{n=0}^{N-1}\binom{A-1-\lfloor n/(b+1)\rfloor}{a-1-\lfloor n/(b+1)\rfloor}.$$

Combining these inequalities and applying the rule $\binom{x}{y-1}+\binom{x}{y}=\binom{x+1}{y}$, we get

$$\sum_{n=0}^{N-1}\left(\sum_{j=0}^{b}\binom{n}{j}\binom{A-n}{a-n+j}\right)\geqslant\sum_{n=0}^{N-1}\binom{A-\lfloor n/(b+1)\rfloor}{a-\lfloor n/(b+1)\rfloor}.$$

Therefore (A, a, b, b) is acceptable. ∎

THEOREM 2. *For any nonnegative integers A, a, B, b such that $A \geqslant a$, $B \geqslant b$, and $\max\{A, B\} \geqslant a+b$, (A, a, B, b) is acceptable.*

Proof. By interchanging A with B and a with b if necessary, we may assume $A \geqslant a+b$ and $B \geqslant b$. We must show that for any N distinct ordered pairs (c_i, d_i) $(i=0, ..., N-1)$, we have the inequality

$$\sum_{n=0}^{N-1}\left(\sum_{j=0}^{b}\binom{n}{j}\binom{A-n}{a-n+j}\binom{B-j}{b-j}\right)-\sum_{n=0}^{N-1}\binom{A-c_n}{a-c_n}\binom{B-d_n}{b-d_n}\geqslant 0.$$

We will show that in fact for any $t_0, ..., t_{b-1} \geqslant 1$,

$$\sum_{n=0}^{N-1} \left(\sum_{j=0}^{b} \binom{n}{j} \binom{A-n}{a-n+j} \prod_{i=j}^{b-1} t_i \right) - \sum_{n=0}^{N-1} \binom{A-c_n}{a-c_n} \prod_{i=d_n}^{b-1} t_i \geqslant 0. \tag{3}$$

The desired inequality will follow upon setting $t_i = (B-i)/(b-i)$.

By Lemma 3, (A, a, b, b) is acceptable; hence (3) holds for $t_0 = \cdots = t_{b-1} = 1$. We now show that (3) still holds when $t_{b-1}, ..., t_0$ are successively increased from 1 to any larger values, which will imply that (3) holds for $t_0, ..., t_{b-1} \geqslant 1$.

Suppose we have established (3) for $t_0 = \cdots = t_m = 1$ and $t_{m+1}, ..., t_{b-1} \geqslant 1$. Viewed as a function of t_m alone, the left side of (3) takes the form $Ct_m + D$, where

$$C \geqslant \left[\sum_{n=0}^{N-1} \left(\sum_{j=0}^{m-1} \binom{n}{j} \binom{A-n}{a-n+j} \right) - \sum_{n=0}^{N-1} \binom{A-c_n}{a-c_n} \right] \prod_{k=m+1}^{b-1} t_k.$$

We have that (A, a, m, m) is acceptable by Lemma 3, which implies that $C \geqslant 0$. Hence if (3) holds for $t_m = 1$, it holds for $t_m \geqslant 1$ as well. ∎

Note that the above proof does not rely on the fact that B is an integer. On the other hand, the inductive proof of Lemma 3 does use the fact that A is an integer; a proof that works for nonintegral A would be a welcome improvement.

3. OBSERVATIONS

Bennett [1] originally considered a variant of the urn-sampling game using two coins in place of the urns, and proved the analogue of Theorem 1 for this variant. He later observed [3] that this analogue could be deduced from the case $-b \leqslant A - B \leqslant a$ of his urn-sampling inequality, using a limiting process in which A and B tend to infinity while a/A and b/B approach a limit. One can think of this as approximating coin tosses by draws from a very large urn.

However, Bennett's results do not apply to the case where only one urn is replaced by a coin. Consequently, he stated the following result as a conjecture, which we now deduce from Theorem 2 by the same limiting argument. For completeness, we give the argument in full.

COROLLARY 1. *The inequality* (2) *holds for the sequences*

$$\mathbf{x}_n = p^n, \qquad \mathbf{y}_n = \binom{B-n}{b}$$

whenever B and b are nonnegative integers with $B \geqslant b$ and $0 < p < 1$.

Proof. Let $\{A_m\}$, $\{a_m\}$ be two sequences of integers such that as $m \to \infty$, $A_m \to \infty$ and $a_m/A_m \to 1-p$. Clearly $A_m \geqslant a_m + b$ for large m. Thus by Theorem 2, the inequality (2) holds for the sequences

$$\mathbf{x}_n = \binom{A_m - n}{a_m}, \qquad \mathbf{y}_n = \binom{B-n}{b}.$$

Since both sides of (2) are linear in $\mathbf{x}$, (2) also holds if $\mathbf{x}_n$ is redefined as

$$\mathbf{x}_n = \frac{\binom{A_m - n}{a_m}}{\binom{A_m}{a_m}} = \prod_{i=0}^{n-1} \frac{A_m - a_m - i}{A_m - i} = \prod_{i=0}^{n-1} \left(1 - \frac{a_m}{A_m - i}\right).$$

Now fix N and take the limit as $m \to \infty$. Then $\mathbf{x}_n \to p^n$ and the desired inequality follows. ∎

One can recover the analogue of Theorem 1 for Bennett's original game by repeating this limiting process. Here we omit the details and simply state the result.

COROLLARY 2. *The inequality* (2) *holds for the sequences*

$$\mathbf{x}_n = p^n, \qquad \mathbf{y}_n = q^n,$$

whenever $0 < p, q < 1$.

A direct combinatorial proof of Corollary 2 was given by Tverberg [4], and a similar proof of Corollary 1 would be of interest. One can first use the argument of the proof of Theorem 2 to reduce to the case $B = b$, in which case the desired inequality is

$$\sum_{n=0}^{N-1} \left(\sum_{j=0}^{b} \binom{n}{j} p^j (1-p)^{n-j} \right) \geqslant \sum_{n=0}^{N-1} (1-p)^{\lfloor n/(b+1) \rfloor}.$$

This inequality should admit a combinatorial interpretation, but the author is currently unaware of one.

Bennett [1] also studied a similar majorization inequality for three or more sequences. While the methods of this paper can readily be generalized to more than two sequences, it appears somewhat complicated to generalize the results of [3] analogously.

ACKNOWLEDGMENT

This work was initiated during a visit to the 1996 Research Experience for Undergraduates at the University of Minnesota, Duluth, directed by Joseph Gallian and sponsored by the National Science Foundation (Grant NSF-DMS 9629078) and the National Security Agency (Grant NSA-MDA 904-96-1-0044).

REFERENCES

1. G. Bennett, Coin tossing and moment sequences, *Discrete Math*. **84** (1990), 111–118.
2. G. Bennett, Double dipping: the case of the missing binomial coefficient identities, *Theoret. Comput. Sci.* **123** (1994), 351–375.
3. G. Bennett, From coin tossing to the Jacobi polynomials, *J. Comb. Theory Series A* **73** (1996), 248–272.
4. H. Tverberg, On a coin tossing problem by G. Bennett, *Discrete Math*. **115** (1993), 293–294.

Printed in Belgium